

高粱 MADS-box 家族基因的鉴定与分析

郑玲, 谢爱玲, 韩建明*

(洛阳师范学院生命科学学院, 河南 洛阳 471934)

摘要: MADS-box 家族是一类具有多种生物功能的转录因子, 它参与了植物生长发育的各个过程。本研究利用生物信息学方法, 从高粱全基因组数据库中筛选并鉴定出 98 个高粱 MADS-box 家族成员, 并对该家族成员的蛋白理化性质、系统进化树、染色体定位、基因结构、基因表达模式等特征进行了全面分析。结果表明 98 个高粱 MADS-box 基因包含了 42 个 M-type-MADS 基因和 56 个 MIKC-MADS 基因。高粱 MADS-box 家族基因的内含子数目为 0~10 个不等, 不同基因的内含子数目相差很大。这些基因位于 1~10 号染色体, 且分布较为均匀。高粱花发育过程中有 28 个 M-type-MADS 基因和 55 个 MIKC-MADS 基因转录表达。果实发育过程中有 32 个 M-type-MADS 基因和 54 个 MIKC-MADS 基因转录表达。

关键词: 高粱; MADS-box 家族; 生物信息学

中图分类号: S514

文献标识码: A

文章编号: 2096-5877(2019)05-0026-04

Identification and Analysis of MADS-box Gene Family of Sorghum Bicolor

ZHENG Ling, XIE Ailing, HAN Jianming*

(College of Life Science, Luoyang Normal School, Luoyang 471934, China)

Abstract: The MADS-box genes family is a class of transcriptional factor with multi-functional, widely involved in many processes during plant growth and development. In this study, a total of 98 MADS-box genes were identified in Sorghum bicolor with bioinformatics methods. And then physico-chemical properties of proteins, phylogenetic relationship, chromosome location, gene structure and gene expression pattern of the MADS-box TFs family in Sorghum bicolor were analyzed. The results showed that the 98 MADS-box genes contained 42 M-type-MADS genes and 56 MIKC-MADS genes. The number of introns in MADS-box family genes varied from 0 to 10. And there were great differences in number of introns in different genes. These genes were evenly distributed on chromosomes 1 to 10. During flower development, 28 M-type-MADS genes and 55 MIKC-MADS genes were expressed. 32 M-type-MADS genes and 54 MIKC-MADS genes were expressed during fruit development.

Key words: Sorghum bicolor; MADS-box family gene; Bioinformatics

高粱 (*Sorghum bicolor*) 是禾本科高粱属植物。目前, 高粱的产量在世界谷物类作物排名第 5^[1]。酿酒酵母的 MCMI、拟南芥的 AGAMOUS、金鱼草的 DEFICIENS 和人类的 SRF4 这四种基因的首字母是 MADS-box 名称的由来。MADS-box 家族基因有一个由 56-58 个氨基酸构成的高度保守区域, 称为 MADS 盒^[2]。根据不同的分类标准, 把 MADS-box 基因分为 Type I 型 (M-type-MADS 型) 和 Type II (MIKC-MADS 型)^[3]。同时又可以 MADS-box 家族基因分成 5 个小类: M α 、M β 、M γ 、MIKC^c、MIKC*, 其中 M α 、M β 、M γ 属于 I 型,

MIKC^c、MIKC* 属于 II 型^[4]。在典型植物中, MADS-box 基因拥有相似的外显子结构和内含子结构, 一般外显子和内含子分别是 7 个、6 个^[5]。通常来说, I 型植物的 MADS-box 基因有 1~2 个外显子, 编码蛋白缺少 K 域但含有 1 个高度保守的 SRF-like MADS 域。II 型植物 MADS-box 基因一般含有 6 个内含子和 7 个外显子^[6]。MIKC 类 MADS-Box 基因在植物基因组中结构和功能的研究相对较为清楚, 它们通常含有 M 域、K 域、I 域和 C 域 4 个结构域, 分别为: MADS-box (M 域)、Kertain-like domain (K 域)、Intervening domain (I 域)、C-terminal domain (C 域)^[7]。MADS-box 家族基因在植物中广泛存在, 它具有调节植物生长发育的作用, 如花发育和果实成熟^[8]。

本研究通过生物信息学的方法鉴定了高粱

收稿日期: 2019-01-27

基金项目: 河南省科技攻关计划 (172102110105)

作者简介: 郑玲 (1981-), 女, 讲师, 硕士, 从事植物学研究。

通讯作者: 韩建明, 男, 博士, 教授, E-mail: 792531425@qq.com

MADS-box 家族基因,并利用各类软件和在线网站进行了全面的分析,这对高粱 MADS-box 家族基因结构和功能的研究具有重要意义。

1 材料和方法

1.1 试验材料

高粱的 MADS-box (SbMADS-box) 转录因子家族数据来自于 NCBI (<http://blast.ncbi.nlm.nih.gov/Blast>)、Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>)、PlantTFDB (<http://plantfdb.cbi.pku.edu.cn/family>) 等数据库。转录因子在高粱花和果实发育中的转录数据来源于 qTeller (<http://qteller.com/>) 网站。其它分析使用的软件以及在线网站包括: InterPro (<http://www.ebi.ac.uk/interpro/>)、MEGA6、GSDS2.0 (<http://gsds.cbi.pku.edu.cn/>)、MapInspect、MEV4。

1.2 试验方法

1.2.1 SbMADS-box 家族基因下载与验证

首先,在 PlantTFDB、Phytozome 数据库下载高粱的 MADS-box 基因序列。在 PlantTFDB 中得到 98 个 MADS-box 基因。在 Phytozome 中下载的 MADS-box 基因有 92 个。两者比对整合,将整合得到的基因在 Interpro 进行验证,确定 98 个基因均为高粱的 MADS-box 家族基因。

1.2.2 SbMADS-box 家族基因进化树构建

把在 PlantTFDB 下载并验证过的 SbMADS-box 基因整合到一个 fas 文件中。同时在 PlantTFDB 中下载拟南芥 (*Arabidopsis thaliana*) MADS-box 基因,将两者用 MEGA6.0 构建进化树并制图。

1.2.3 基因结构分析

利用 Phytozome 下载高粱的 MADS-box 基因家族成员的 cDNA 序列以及 CDS 序列,使用 GSDS 在线分析结果并制图,绘制出高粱 MADS-box 基因的内含子-外显子基因结构图。

1.2.4 SbMADS-box 家族基因染色体定位分析

首先在 Phytozome 中得到每个基因序列的染色体位置。将位置信息整理,同一染色体的基因排列顺序要求从小到大排列,单位转换为“mp”,每条染色体的基因数用“nloc”表示,染色体序号用“group”表示,不明确所在染色体位置的基因用 N/A 表示。将整理的结果利用 MapInspect 制成染色体定位图片。

1.2.5 SbMADS-box 家族基因花、果实发育过程表达模式分析

利用 qTeller 网站中的高粱转录组数据库,找

到 SbMADS-box 家族基因在花、果实发育过程中的表达数据,并利用 MEV4 软件对检测到的数据进行分析,制作热图。

2 结果与分析

2.1 SbMADS-box 家族基因基本信息

在 PlantTFDB 网站上得到蛋白质的基因长度(aa)、理论等电点(PI)、蛋白分子量 MW(Da)等基本信息。根据高粱 MADS-box 家族基因在染色体上基因的位置将 98 个基因重新命名。用缩写 Sb 代表高粱,根据基因的位置信息及其先后顺序对 98 个高粱 MADS-box 家族基因进行命名,其中 42 个 M-type-MADS 家族基因命名为 SbMADS-box1 ~ SbMADS-box42, 56 个 MIKC-MADS 基因命名为 SbMADS-box43 ~ SbMADS-box98。

2.2 SbMADS-box 家族基因进化树的构建

利用邻接法 (Neighbor-joining, NJ)^[1] 在软件 MEGA6.0 构建高粱和拟南芥 MADS-box 家族基因的进化树。根据高粱和拟南芥家族基因的系统进化树,将进化树进行亚族划分。根据拟南芥的亚族分类来划分 MADS-box 家族。42 个 M-type-MADS 基因可分为 M α 、M β 、M γ 亚族,56 个 MIKC-MADS 基因可分为 MIKC^c、MIKC* 亚族。

2.3 SbMADS-box 家族基因结构分析

结果显示 98 个 MADS-box 家族基因中有 29 个基因不含内含子,其余 69 个 MADS-box 家族基因分别含有 1 ~ 10 个内含子不等。在进化树上同一分支的基因内含子数目较为相似。

42 个 M-type-MADS 家族基因中,有 28 个基因不含内含子,大约占据了全部基因的 2/3。剩下的基因中有 7 个基因只有一个内含子;只有 SbMADS-box32 有 2 个内含子;3 个基因有 4 个内含子;只有 SbMADS-box34 有 5 个内含子;SbMADS-Box27 和 SbMADS-box28 有 10 个内含子。

56 个 MIKC-MADS 家族基因中,只有 SbMADS-box54 不含内含子,其余基因中有 3 个内含子的有 SbMADS-box59、SbMADS-box96 和 SbMADS-box97;有 4 个内含子的基因是 SbMADS-Box58;有 5 个内含子的基因是 SbMADS-box63、SbMADS-box78、SbMADS-box87 和 SbMADS-box92;SbMADS-box95 有 10 个内含子;剩余的 46 个基因有 5 个或者 6 个内含子。

2.4 SbMADS-box 家族基因的染色体定位

高粱 98 个 MADS-box 基因中,56 个 MIKC-MADS 基因和 41 个 M-type-MADS 基因都分布在

1~10号染色体上, SbMADS-box42目前还不能确定其位于哪条染色体上, 用N/A(unanchored)表示。

从染色体定位结果来看, 在高粱的M-type-MADS家族基因中41个基因均匀分布在第1、2、3、4、5、6、8、9、10号染色体上, 分别含有2个、6个、8个、2个、10个、5个、1个、3个、4个基因。6号染色体没有M-type-MADS基因。1、2、5、9号染色体上的基因较为集中分布, 2、3、6、10号染色体上较均匀分布, 但局部也存在成群分布的现象。

在高粱的MIKC-MADS家族基因中, 56个基因全部分布在第1、2、3、4、5、7、8、9、10号染色体上, 分别含有10个、4个、5个、11个、6个、3个、6个、5个、6个基因。6号染色体没有MIKC-MADS基因。基因在各个染色体上的分布比较松散, 但从单一的染色体来看, 又较为集中。第1、4、5、8、9号染色体均有局部集群的特点。而其他第2、3、7、10号染色体呈现相对分散的状态。

2.5 SbMADS-box家族基因在高粱花和果实发育过程中的表达

2.5.1 SbMADS-box家族基因在高粱花发育过程中的表达

42个M-type-MADS基因中有28个基因在高粱花发育中进行了表达。56个MIKC-MADS基因中有55个基因进行了表达。

对下载到的高粱叶(leave)、花药(anther)、花序(inflorescence)、雌蕊(pistil)发育过程中M-type-MADS基因的转录量进行分析:(1)叶发育过程中SbMADS-box1/2/10/13/14/17/18/19/20/21/22/23/24/25/26/27/28/33/37/38/39这21个基因转录量较高, 说明这些基因在叶的发育过程中有重要作用。(2)花药发育过程中SbMADS-box1/13/14/18/19/20/21/22/23/24/25/26/35/38/39这15个基因具有高转录量。(3)在早期花序发育中, 转录量较高的11个基因是SbMADS-box1/3/7/17/18/19/21/23/25/35/37。(4)新生花序发育具有高转录量的15个基因是SbMADS-box1/3/7/16/17/18/19/20/21/22/23/24/25/26/41。(5)雌蕊发育中有11个基因转录量较高, SbMADS-box1/3/7/10/20/22/23/24/25/26/35, 说明这些基因和雌蕊的发育密切相关。

MIKC-MADS基因在花发育过程中的转录表达特点为:(1)在叶发育中SbMADS-box56/58/66/67/68/69/79/81/82/83/84/85/92这13个基因转录量较高, 说明这些基因在叶的发育过程中有重要作用。(2)SbMADS-box58/64/65/66/67/68/69/70/82/83/

84/85/88/89/90/91/94这17个基因具有高转录量, 说明这些基因在花药发育方面有决定性作用。(3)在早期花序发育中, 转录量较高的15个基因是SbMADS-box57/59/70/71/72/82/83/84/85/86/88/89/90/91/94。(4)新生花序发育过程中转录量较高的15个基因是SbMADS-box59/61/63/66/67/68/69/70/71/78/82/83/84/85/95。(5)雌蕊发育中有9个基因具有高转录量, SbMADS-box71/72/77/78/82/83/84/85/93, 这些基因和雌蕊的发育密切相关。

2.5.2 SbMADS-box家族基因在高粱果实发育过程中的表达

果实发育的数据以及热图显示, 有86个基因在高粱种子(seed)、胚(Embryo)、胚乳(Endosperm)发育中进行了表达, 其中包括32个M-type-MADS基因和54个MIKC-MADS基因。

对下载到的高粱种子、胚、胚乳发育过程中M-type-MADS基因的转录量进行分析:(1)在种子发育过程中SbMADS-box7/15/19/21/33/34/38/40/41转录量较高, 说明这些基因在种子的发育过程中有重要作用。(2)在胚发育过程中, SbMADS-box15/16/40转录量较高, 即它们在胚发育过程中作用显著。(3)在胚乳发育过程中, SbMADS-box15/33/40转录量较高, 说明这些基因在胚乳发育中有决定性作用。

MIKC-MADS基因在果实发育过程中的转录表达特点为: 在种子发育过程中SbMADS-box43/44/48/52/55/56/57/60/61/62/63/70/77/78/86/87/88/89/90/91/92转录量较高, 说明这些基因在种子的发育过程中有重要作用。在胚发育过程中, SbMADS-Box58/62具有高转录量, 即它们的作用较为显著。在胚乳发育过程中, SbMADS-box48/55/57/61/70/71/72/81/86/87/88/89/90/91/92/98转录量较高, 说明这些基因在胚乳发育中有决定性作用。

3 讨论

目前国内外学者已经对拟南芥^[9]、水稻^[10]、芝麻^[11]、葡萄^[12]和油菜^[13]等多个物种的MADS-box基因家族进行了研究。最近几年的研究显示MADS-box家族基因在高等植物花发育的过程中发挥着重要作用。芸薹属植物的MADS-box基因与花器官分化^[14]、开花时间^[15]相关。银杏MADS-box基因在银杏开花调控中发挥重要作用, 是花器官发育过程中的关键基因^[16]。建兰MADS-box基因在蕊柱的形成过程中可能具有关键角色^[17]。MADS-box基因在兰花的花器官形成过程中起重

要作用,A、B、C、D和E类基因共同作用形成独特及多样的兰花结构^[18]。桑树MADS-box基因可能与内休眠过程相关^[19]。

高粱是干旱和半干旱地区广泛种植的一种粮食、饲料和能源作物^[20],是世界上仅次于玉米、小麦、水稻和大麦的第5大粮食作物,具有耐旱、耐涝、耐贫瘠、耐盐碱等多重抗性^[21]。本研究共鉴定和分析了98个高粱MADS-box家族基因,包括了42个M-type-MADS基因(I型MADS-box基因)和56个MIKC-MADS基因(II型MADS-box基因)。M-type-MADS基因可分为M α 、M β 、M γ 亚族,MIKC-MADS基因可分为MIKC^c、MIKC*亚族。M-type-MADS基因和MIKC-MADS基因的内含子数目都在0~10个之间。M-type-MADS基因中不含内含子的基因数目大约占全部I型基因的2/3。MIKC-MADS基因家族中不含内含子的基因只有一个,且含有6个内含子或者7个内含子的基因的占全部II型基因的80%还要多。

98个基因中有97个基因在相应染色体上有确切位置,只有SbMIKC-MADS42的染色体位置尚未明确。这些基因位于1~10号染色体,且分布较为均匀。

高粱花发育过程中的转录表达数据显示,有83个基因在高粱花药、花序、雌蕊发育中进行了表达,其中包括28个M-type-MADS基因和55个MIKC-MADS基因。高粱果实发育过程的转录表达数据显示,有86个基因在高粱种子、胚、胚乳发育中进行了表达,其中包括32个M-type-MADS基因和54个MIKC-MADS基因。

本研究对高粱MADS-box基因家族进行了初步分析,为深入研究该基因家族的基因结构和功能等提供参考数据,同时也为弄清高粱花和果实发育过程中MADS-box转录因子调控机制奠定了基础。

参考文献:

- [1] 孙璐.高粱耐盐品种筛选及耐盐机制研究[D].沈阳:沈阳农业大学,2012.
- [2] 胡丽芳,金志强,徐碧玉,等.MADS-box基因在果实发育、成熟过程中的作用[J].分子植物育种,2005,3(3):415-420.
- [3] Theissen G, Becker A, Di Rosa A, et al. A short history of MADS-box genes in plants[J]. Plant Molecular Biology,2000,42(1):115-149.
- [4] 朱友银,邵娟,王月,等.樱桃MADS-box转录因子的生物信息学及其表达分析[J].植物生理学报,2015,51(3):354-362.
- [5] 张远森,秦晓萌,严金平,等.香蕉MADS-box基因家族的生物学信息分析[J].植物生理学报,2013,49(10):1030-1040.
- [6] 黄方,迟英俊,喻德跃.植物MADS-box基因研究进展[J].南京农业大学学报,2012,35(5):9-18.
- [7] Ogata K, Kanei-Ishii C, Sasaki M, et al. The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation[J]. Nature structural biology, 1996,3(2):178-187.
- [8] Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in Arabidopsis thaliana[J]. Current Opinion in Plant Biology, 2001,4(5):447-456.
- [9] Parenicová L, de Folter S, Kieffer M, et al. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world[J]. Plant Cell, 2003, 15: 1538-1551.
- [10] Arora R, Agarwal P, Ray S, et al. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress[J]. BMC Genomics, 2007, 8:242.
- [11] Wei X, Wang L H, Yu J Y, et al. Genome-wide identification and analysis of the MADS-box gene family in sesame[J]. Gene, 2015, 569: 66-76.
- [12] Diaz-Riquelme J, Lijavetzky D, Martínez-Zapater J M, et al. Genome-wide analysis of MIKC^c-type MADS-box genes in grapevine[J]. Plant Physiol, 2009,149: 354 - 369.
- [13] 高虎虎,张云霄,胡胜武,等.甘蓝型油菜MADS-box基因家族的鉴定与系统进化分析[J].植物学报,2017,52(6):699-712.
- [14] Fang S C, Fernandez D E. Effect of regulated over-expression of the MADS domain factor AGL15 on flower senescence and fruit maturation[J]. Plant Physiol. 2002,130:78 - 89.
- [15] Chang Y Y, Chiu Y F, Wu J W, et al. Four orchid (Oncidium Gower Ramsey) AP1/AGL9-like MADS-box genes show novel expression patterns and cause different effects on floral transition and formation in Arabidopsis thaliana[J]. Plant Cell Physiol, 2009, 50: 1425-1438.
- [16] 董金金,刘伟,李萌,等.银杏MADS-box基因家族的表达及系统发育分析[J].植物生理学报,2018,54(6):1055-1063.
- [17] 向林,陈跃,陈丽萍,等.建兰花发育相关B、C和E类MADS-box基因的表达分析[J].园艺学报,2018,55(8):1595-1604.
- [18] 胡月苗,向林章,秋爽,等.兰科植物花器官发育MADS-box调控基因研究进展[J].分子植物育种,2018(4):886-895.
- [19] 罗义维.桑树休眠相关MADS-box基因的功能研究[D].重庆:西南大学,2018.
- [20] 刘方明,高玉山,孙云云,等.高粱抗旱性鉴定研究进展[J].东北农业科学,2016,41(3):5-7.
- [21] 张春宵,王晶,王风华,等.分子标记在高粱研究中的应用[J].吉林农业科学,2013,38(2):42-46.